Military Psychology
2014, Vol. 26, No. 3, 131–137

0899-5605/14/$12.00 http://dx.doi.org/10.1037/mil0000041

Introduction to the Special Issue on Selected New Developments in Military Enlistment Testing

Michael G. Rumsey
Annandale, Virginia

Jane M. Arabian
Office of the Under Secretary of Defense for
Personnel and Readiness,
Washington, DC

This article provides a brief discussion of the historical context for the major recent advances in enlisted selection and classification that are the subject of this special issue. By providing a perspective on the difficulty personality assessment has had gaining traction as a screening measure, it supplies a basis for understanding the significance of recent advances in personality measurement. This article also summarizes the military research on interest measurement that will receive more extended treatment later in this issue. It relates the work in this issue on new information technology testing to the role of information testing in the current military operational selection and classification test battery, the Armed Services Vocational Aptitude Battery (ASVAB), and discusses the background for 2 additional cognitive measures, Coding Speed and Assembling Objects. Finally, it previews issues to be discussed in the commentary article at the end of this issue.

Keywords: selection, classification, personality, Armed Services Vocational Aptitude Battery, interests

Since 1917, with the development of the Army Alpha and Beta, the military services have been at the forefront of selection and classification research and practice. The history of military enlistment testing has been marked by several important developments since the implementation of these early tests, including the gradual evolution from measures of general cognitive ability to more differentiated multiple aptitude measures designed to serve classification as well as selection, the origination of a single joint-service battery, the Armed Services Vocational Aptitude Battery (ASVAB) to replace those developed separately by the Army, Navy, and Air Force, the transition from an exclusively paper and pencil format to the inclusion of a computerized adaptive format, the validation of the joint-service battery against performance in each of the services, and the addition of a spatial test to this battery (Defense Manpower Data Center, 2012; Defense Manpower Data Center, n.d.; Rumsey, 2012).

As the 20th century gave way to the 21st, a new revolution in military testing could be observed. Noncognitive tests, which had previously played at best a secondary role in such testing, were now a matter of keen interest. New content was envisioned, especially with respect to the increased focus on information technology in military jobs. The need to reconsider existing content and testing methodology also came into focus. These themes were given prominence by an ASVAB Review Panel, established to provide recommendations concerning military testing (Drasgow, Embretson, Kyllonen, & Schmitt, 2006).

Complete treatment of all these changes is beyond the scope of this special issue. This volume highlights new noncognitive tests and their implementation in multiple screening contexts, a new information and communications technology literacy test, a perceptual

---

Michael G. Rumsey, Annandale, Virginia; Jane M. Arabian, Office of the Under Secretary of Defense for Personnel and Readiness, Washington, DC.

The opinions expressed are those of the authors and not necessarily those of the United States Government, Department of Defense.

Correspondence concerning this article should be addressed to Michael G. Rumsey, Annandale, VA 22003. E-mail: miker1998@aol.com

speed and accuracy test, and a spatial ability test.

Noncognitive Tests

Two types of noncognitive measures have received particular attention in the history of military enlistment testing. One type might be referred to as personality, although measurement of personality constructs may be indirect and even unintentional. For example, background and interest measures may reveal aspects of an individual's personality even if the relevant personality constructs are not specified. The terms personality and temperament are often used interchangeably in the military literature. For a time, the term "personality" was avoided in the context of Army selection and classification testing research to preclude confusion with more clinically oriented measures often identified as personality tests.

The second type is vocational interests, defined by Van Iddekinge, Putka, and Campbell (2011, p. 14) as "relatively stable individual differences that influence behavior through preferences for certain work activities and work environments." The distinction between vocational interest testing and personality testing is often difficult to draw, as interests may be either a reflection of an individual's personality or vocational preference. Some tests may have elements of both types. However, in recent decades the distinction has been more consciously drawn, so we will treat the types separately even while acknowledging the overlap.

Personality Testing

Although interest in noncognitive correlates of officer performance can be traced to the use of situation tests in Germany as early as 1927 (Rumsey & White, 2010), little attention was given to personality testing for enlistment selection before the 1950s. Then, measures that "captured aspects of personality such as self-confidence, self-assertiveness, emotional stability, leadership, and devotion to group objectives" (Greenston, 2012) were added to the Army battery used for classifying soldiers into jobs, the Army Classification Battery. Each of the services developed background measures for enlisted assessment in the 1970s—the Air Force's History Opinion Inventory (HOI), the Navy's Recruiting Background Questionnaire (RBQ), and the Army's Military Applicant Profile (MAP) (Rumsey & White, 2010). However, when the services combined their efforts to implement the joint-service ASVAB in 1976, it contained only cognitive measures (Rumsey, 2012).

The value of personality measures for selection has been questioned during much of the history of standardized testing. A review of available research (Guion & Gottier, 1965) that summarily dismissed the value of personality measures in personnel selection reflected the general attitude toward such measures at the time. A more promising result was obtained with the large scale Army selection and classification project initiated in the 1980s known as Project A. Using a thoughtful and thorough development process involving the identification of potentially relevant personality dimensions through a comprehensive literature review, an expert judgment task focused on expected validity of each dimension (Peterson & Wing, 2001), and large-scale pretesting, researchers constructed an instrument known as the Assessment of Background and Life Experiences (ABLE; Hough, Barge, & Kamp, 2001). The validity of the ABLE was determined based on two types of performance criteria—(a) task proficiency ("can do" performance) and (b) a combination of ratings and administrative measures that were determined to represent the extent to which the individual performed over time ("will do" performance). The ABLE added little to the prediction of proficiency, but contributed much to the prediction of the will do criteria (Oppler, McCloy, Peterson, Russell, & Campbell, 2001).

ABLE was also found to be an effective predictor of premature separation from the Army (i.e., attrition) (White, Young, & Rumsey, 2001). A joint-service measure drawn from previous research by each of the services, the Armed Services Applicant Profile (ASAP), was also found to be an effective predictor of military attrition (Trent, 1993). The Department of Defense (DoD) saw the potential for use of a measure that combined the ABLE and ASAP for enlisted screening, the Adaptability Screening Profile (ASP), but implementation was thwarted by a concern that when used operationally, applicants' motivation to present themselves positively rather than accurately, often pejoratively and unsympathetically referred to as "faking," would undermine the utility of the measure (Trent, 1993).

Strategies to control faking then became a major focus of military research on enlistment personality tests. Fan et al. (2012) identified three paradigms for faking mitigation—reactive, preventative, and a combination of the first two. The Army's effort to develop a faking resistant attrition screen focused on prevention, through a multidimensional forced choice strategy. The rationale for this approach, the manner in which it was translated into a measurement instrument, how the Army perceived its screening potential, how it was combined with other measures to form what became known as the Tier Two Attrition Screen (TTAS), and further developments in attrition screening, constitute the content of the next article in this special issue, by White, Rumsey, Mullins, Nye, and LaPort (2014, pp. 138–152).

The Army's success in developing and implementing a personality screen for the prediction of attrition stimulated interest in wider use of personality measures in the enlistment screening process. The TTAS was implemented only for screening nonhigh school diploma graduates, who by policy constitute only a small portion of military enlistments (Young & White, 2006). Why not, Army leaders asked, use personality measures in screening at least a portion of the high school diploma graduates? This interest was accompanied by a new development in personality measurement. A group of researchers from the Drasgow Consulting Group (DCG) were exploring the viability of computer adaptive personality testing (e.g., Stark, Chernyshenko, & Drasgow, 2005). Computer adaptive testing could offer efficiencies that paper and pencil testing could not match, but the approach as applied to personality measurement presented such conceptual and psychometric challenges that its adoption had hardly been contemplated before the DCG researchers began their work. After the article on attrition screening, Stark et al., (2014, pp. 153–164) will describe how the DCG work for the Army led to the development and experimental implementation of the Tier One Performance Screen, and how a separate DCG effort resulted in an experimental Navy computer adaptive screening instrument.

Vocational Interest Testing

Vocational interest testing focuses on the fit between individual interests and the environment and tasks associated with a particular job, as well as the congruence of interests between one potential or actual job incumbent and the other incumbents. The premise behind such testing is that the better the fit, the greater the likelihood that the individual will be satisfied with the job, leading to greater retention and possibly even better performance.

The first significant work on an interest measure designed for enlisted military applicants was the Navy's development of the Navy Vocational Interest Inventory (NVII), a process that began in 1946 and was completed in 1961. Existing commercial instruments, designed for those seeking careers requiring a college education, were deemed inappropriate for those who would be seeking the types of blue collar jobs available to enlisted service members. Thus, the NVII was built from the ground up, based on activities that reflected divergent interests between members in different Navy occupations (Clark, 1955, 1961). Validation research indicated that, when the jobs of Navy enlistees were aligned with their interests, these enlistees were indeed more likely to reenlist and perform better (Lau, 1969).

The Air Force followed with the development of the Vocational Interest Career Examination (VOICE) in the early 1980s. Like the Navy, the Air Force eschewed existing commercial instruments, developing items linked to Air Force jobs. The VOICE consisted of four types of items: occupational titles, work tasks, leisure time activities, and desired learning experiences. Respondents indicated preferences in terms of like, indifferent, or dislike. The measure was found to be a good predictor of job satisfaction (Alley & Matthews, 1982).

A few years later, when the Army desired to test a vocational interest inventory in Project A, it used the VOICE as a starting point. The Army's version, named the Army Vocational Interest Career Examination (AVOICE), was refined and revised to meet Army needs. The AVOICE has been found to predict performance and provide results that are consistent with one's current military occupational specialty (MOS) (Hanson, Paullin, Bruskiewicz, & White, 2003).

The DoD also initiated a program of interest assessment as part of its Career Exploration Program (CEP). This program, initiated in 1992, involves administration of both the ASVAB and an interest measure to high school and college students to both provide information to help them explore career options and to identify qualified leads for service recruiters. Initially, the interest inventory used in the CEP was one already available, the Self Directed Search (SDS; Holland, 1985). This instrument was based on Holland's vocational personality theory, which determined that person-job match could be based on six interest categories collectively named RIASEC: Realistic (R), Investigative (I), Artistic (A), Enterprise (E), and Conventional (C) (Wall, Wise, & Baker, 1996).

The SDS was then used to validate the DoD instrument, the Interest-Finder, in the CEP. A revised interest inventory, Find Your Interests, is used in the current ASVAB CEP. DoD decided to develop its own instrument to better focus on the youth population it was testing. Although it continued to use the RIASEC model, it defined its own content areas and types of items for the Interest-Finder, which was introduced in 1995 (Wall, Wise, & Baker, 1996). In recent years, the Army has been experimenting with its own RIASEC-based measure of interest in a variety of work activities, the Work Preferences Survey (WPS), which has been found to relate well to a variety of performance and attitudinal criteria (Putka & Van Iddekinge, 2007; Van Iddekinge et al., 2011).

For several years, the Navy has also been developing a new computerized interest inventory, known as Jobs and Occupational Interest in the Navy (JOIN). Components of this inventory include work areas, work styles, work environments, and work activities. A key characteristic of this inventory involves the use of pictures to represent elements of each component (Farmer, Watson, Alderton, Michael, & Hindelang, 2006). Ingerick and Rumsey (2014, pp. 165–181) describe these developments in interest measurement for those in an enlistment-eligible population in a separate article in this special issue.

Cognitive Testing

Information Technology

In the earliest days of military testing, the focus was on general aptitude, not on specific capabilities that might translate to differential performance across jobs. However, as the credibility of military testing grew, so did the interest in building tests that could be effective in job placement. As the services began to build classification batteries, they focused particularly on information tests that might indicate an individual's interest and aptitude with respect to particular subject areas. The current ASVAB has several information tests designed principally for use in person-job match. As the ability to receive, transmit, and interpret computerized information began to play an increasingly greater role in the success of military operations, the lack of a measure specifically focused on such ability has become a major concern among the services. The Air Force was given the lead on developing such a measure, with the other services also participating in its validation. Therefore, another article in this special issue, by Trippe, Moriarty, Russell, Carretta, and Beatty (2014, pp. 182–198) will deal with the research that has been conducted on this instrument.

Perceptual Speed and Accuracy

Before 2002, the ASVAB contained two speeded tests, Numerical Operations and Coding Speed. These tests provided coverage of one of the factors identified in the ASVAB, Clerical Speed (Alderton, Wolfe, & Larson, 1997), and aided in classification. However, it was difficult to maintain scoring consistency for these tests across diverse testing conditions, and they were discontinued as Joint-Service tests. The Navy alone continued to administer Coding Speed as a special test and has amassed evidence supporting its reintroduction as part of the Joint-Service testing battery. This research is discussed in this issue by Held, Carretta, and Rumsey (2014, pp. 199–220).

Spatial Ability

The ASVAB originally had a measure of spatial ability, Space Perception (Maier, 1993), although this test was later dropped. In Project A, spatial tests were found to provide incremental validity over the ASVAB in predicting task performance in Army jobs. Further supportive evidence of the value of spatial tests in predicting job performance in the military led to the decision to include a measure of spatial ability, Assembling Objects, in the ASVAB. The article by Held and her co-authors referred to above discusses research recently conducted supporting the use of Assembling Objects for enlisted classification.

Issues and Implications

The research described in this special issue carries with it a number of concerns and implications, which will be discussed in the Commentary, the final article. Perhaps the most significant issue is whether this research is on the right track, with respect to such questions as:

1. Is it addressing those attributes that, beyond those already being measured, are most important for success in the military environment? Does the research indicate that an approach combining cognitive and noncognitive measures is likely to provide greater benefits to the military than one focused only on cognitive aptitudes? Is the research focusing on the right noncognitive attributes and is it measuring them in the most effective way? With respect to cognitive testing, is there any benefit to be gained by adding another information or specific aptitude test? Will this just amount to another measure of general aptitude that is so highly correlated with the current operational ASVAB tests that it adds nothing, or will it identify those who, through their understanding and inclination, are uniquely qualified to perform vital tasks in certain jobs?
2. Is it using the most appropriate criteria to determine effectiveness? Historically, two types of outcomes have been of greatest interest in military selection and classification research: retention (defined here as continuation both within and beyond one's initial obligated tour of service) and performance. Project A demonstrated that there are two principal classes of performance that must both be considered if one is to have a complete picture of an individual's benefit to the organization. One has to do with job proficiency, the other with the motivation to perform effectively over time (Rumsey, Peterson, Oppler, & Campbell, 1996). In the commentary, the information provided by the criteria used in each investigation will be examined.

The commentary will also consider the contributions made by this research to military selection and classification and to psychological research in general. It will examine the extent to which new measures described in this special issue have been implemented and the lessons that can be drawn from this recent history. It will also examine how these research activities have contributed to the science of attribute measurement for selection and classification. Military testing has always been both a consumer of, and contributor to, the broader field of personnel testing.

Finally, the commentary will consider issues for future research. It will consider neglected areas in the realm of individual differences that might be primary candidates for future research. It will suggest that pessimism regarding the potential classification gains possible with existing measures is misplaced, but will note that the lack of available cost-effective performance measures is a particular handicap for research needed to (a) assess continuously the value of current measures as job requirements and the operational military environment changes, and (b) examine the potential value of new measures for classification purposes. Thus, research to both improve performance measurement and to enhance tools for estimating validity for jobs lacking performance measures should be important future priorities.

References

Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT battery. *Military Psychology, 9,* 5–37.

Alley, W. E., & Matthews, M. D. (1982). The Vocational Interest Career Examination: A description of the instrument and possible applications. *Journal of Psychology: Interdisciplinary and Applied, 112,* 169–193. doi:10.1080/00223980.1982.9915374

Clark, K. E. (1955). *The use of interest measures with naval enlisted personnel*. Minneapolis, MN: University of Minnesota Press.

Clark, K. E. (1961). *The vocational interests of nonprofessional men*. Minneapolis, MN: University of Minnesota Press.

Defense Manpower Data Center. (2012). *P&P ASVAB forms 23–27* (Tech. Bulletin No. 4). Seaside, CA: Defense Manpower Data Center.

Defense Manpower Data Center. (n.d.). *History of military testing*. Retrieved from http://official-asvab.com/history_res.htm

Drasgow, F., Embretson, S. E., Kyllonen, P. C., & Schmitt, N. (2006). *Technical review of the Armed Services Vocational Aptitude Battery* (Final Rep. No. 06–25). Alexandria, VA: Human Resources Research Organization.

Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology, 97*, 866–880. doi:10.1037/a0026655

Farmer, W. L., Watson, S. E., Alderton, D. L., Michael, P. G., & Hindelang, R. (2006). *Improving person-job congruence during the classification process: Item development and initial testing of a pictorial interest instrument* (Tech. Note No. 06–8). Millington, TN: Navy Personnel Research, Studies and Technology Division, Bureau of Naval Personnel.

Greenston, P. (2012). Classification research. In P. F. Ramsberger, N. R. Wooten, & M. G. Rumsey (Eds.), *A history of the research into methods for selecting and classifying U.S. Army personnel, 1917–2010* (pp. 275–338). Lewiston, NY: Edwin Mellen Press.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18,* 135–164. doi:10.1111/j.1744-6570.1965.tb00273.x

Hanson, M. A., Paullin, C. J., Bruskiewicz, K. T., & White, L. A. (2003, November). *The Army Vocational Interest Inventory*. Paper presented at the 45th Annual Conference of the International Military Testing Association, Pensacola, FL.

Held, J. D., Carretta, T. R., & Rumsey, M. G. (2014). Evaluation of tests of perceptual speed/accuracy and spatial ability for use in military occupational classification. *Military Psychology, 26,* 199–220. doi:10.1037/mil0000043

Holland, J. L. (1985). *Making vocational choices: A theory of vocational personalities and work environments* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Hough, L., Barge, B., & Kamp, J. (2001). Assessment of personality, temperament, vocational interests, and work outcome preferences. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 111–154). Mahwah, NJ: Erlbaum.

Ingerick, M., & Rumsey, M. G. (2014). Taking the measure of work interests: Past, present, and future. *Military Psychology, 26,* 165–181. doi:10.1037/mil0000045

Lau, A. W. (1969, September). *The Navy Vocational Interest Inventory as a predictor of school performance, job performance, and reenlistment*. Paper presented at the 11th Annual Conference of the Military Testing Association, New York.

Maier, M. H. (1993). *Military aptitude testing: The past fifty years* (Tech. Rep. No. 93–007). Monterey, CA: Personnel Testing Division, Defense Manpower Data Center.

Oppler, S. H., McCloy, R. A., Peterson, N. G., Russell, T. L., & Campbell, J. P. (2001). The prediction of multiple components of entry-level performance. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 389–409). Hillsdale, NJ: Erlbaum.

Peterson, N. G., & Wing, H. (2001). The search for new measures: Sampling from a population of selection/classification predictor variables. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits of personnel selection and classification* (pp. 53–69). Mahwah, NJ: Erlbaum.

Putka, D. J., & Van Iddekinge, C. H. (2007). Work preferences survey. In D. J. Knapp & T. R. Tremble (Eds.), *Concurrent validation of experimental Army enlisted personnel selection and classification measures* (Tech. Rep. No. 1205). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Rumsey, M. G. (2012). Military selection and classification in the United States. In J. H. Laurence & M. D. Matthews (Eds.), *The Oxford Handbook of Military Psychology* (pp. 129–147). New York, NY: Oxford University Press.

Rumsey, M. G., Peterson, N. G., Oppler, S. H., & Campbell, J. P. (1996). What's happened since Project A: The future career force. *Journal of the Washington Academy of Sciences, 84,* 94–110.

Rumsey, M. G., & White, L. A. (2010). Benefits and barriers to personality measures in military personnel selection. In P. T. Bartone, B. H. Johnsen, J. Eid, J. M. Violanti, & J. C. Laberg (Eds.), *Enhancing human performance in security operations* (pp. 85–119). Springfield, IL: Charles C Thomas.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-dimensional pairwise-preference model. *Applied Psychological Measurement, 29,* 184–203. doi:10.1177/0146621604273988

Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions. *Military Psychology, 26,* 153–164. doi:10.1037/mil0000044

Trent, T. (1993). The Armed Services Applicant Profile (ASAP). In T. Trent & J. H. Laurence (Eds.), *Adaptability screening for the Armed Forces* (pp. 41–70). Washington, DC: Office of Assistant Secretary of Defense (Force Management and Personnel).

Trippe, D. M., Moriarty, K. O., Russell, T. L., Carretta, T. R., & Beatty, A. S. (2014). Development of a cyber/information technology knowledge test for military enlisted technical training qualification. *Military Psychology, 26,* 182–198. doi:10.1037/mil0000042

Van Iddekinge, C. H., Putka, D., & Campbell, J. P. (2011). Reconsidering vocational interests for personnel selection: The validity of an interest-based selection test in relation to job knowledge, job performance, and continuance intentions. *Journal of Applied Psychology, 96,* 13–33. doi:10.1037/a0021193

Wall, J. E., Wise, L. L., & Baker, H. E. (1996). Development of the Interest Finder: A new RIASEC-based interest inventory. *Measurement and Evaluation in Counseling and Development, 29,* 134–152.

White, L. A., Rumsey, M. G., Mullins, H. M., Nye, C. D., & LaPort, K. A. (2014). Toward a new attrition screening paradigm: Latest Army advances. *Military Psychology, 26,* 138–152. doi:10.1037/mil0000047

White, L. A., Young, M. C., & Rumsey, M. G. (2001). Assessment of Background and Life Experiences (ABLE) implementation issues and related research. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification* (pp. 525–558). Hillsdale, NJ: Erlbaum.

Young, M. C., & White, L. A. (2006). *Preliminary operational findings from the Army's Tier Two Attrition Screen (TTAS*). Paper presented at the annual meeting of the American Psychological Association, Honolulu, HI.

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/authors/review-manuscript-ce-video.aspx.